Semistructured Temporal clinical geographical Systems

http://stars.di.univr.it
1. The research group
2. Courses
   - Bachelor courses
   - Master courses
3. Research
   - Research activities
   - Projects
   - Publications
4. Theses and more
1. The research group

2. Courses
   - Bachelor courses
   - Master courses

3. Research
   - Research activities
   - Projects
   - Publications

4. Theses and more
Who we are

- Carlo Combi
  Full Professor
- Alberto Belussi
  Associate Professor
- Damiano Carra
  Assistant Professor
- Barbara Oliboni
  Assistant Professor
- Roberto Posenato
  Assistant Professor
- Pietro Sala
  Post-Doc
- Sara Migliorini
  Post-Doc
- Gabriele Pozzani
  Post-Doc
- Alberto Sabaini
  Ph.D. Student
What we do

The group focuses on the

- theoretical
- technological
- methodological

aspects of the information management in complex organizations

The research activities comprise:

- temporal databases
- spatial databases
- semistructured databases
- information systems on the web
- temporal information systems
- information systems in medicine
- geographical information systems
- process-aware information systems
The STARS lab is located at the floor -2 in Ca’ Vignal 2. It’s a lab where students collaborating with the group can develop their projects and theses.

The lab has 10 workspaces and includes 5 PCs equipped with Ubuntu and any software useful for the research and thesis work, e.g., PostgreSQL, MySQL, Java SDK, Pentaho.

The PCs in the lab are connected with the STARS server that manages the users’ authentication and disk space.
Outline

1. The research group
2. Courses
   - Bachelor courses
   - Master courses
3. Research
   - Research activities
   - Projects
   - Publications
4. Theses and more
Provided courses

People in the research group give the following courses:

- in bachelor programs:
  - Databases
  - Databases for bioinformatics
  - Networks
  - Software Engineering, Lab

- in master programs:
  - Information systems in medicine
  - Advanced databases
  - Geographic and multimedia information systems
  - Data-intensive computing systems
  - Biomedical and bioinformatics databases
  - Information systems
  - Complexity
  - Information retrieval
The course focuses on two main arguments:

- the design of a database and its applications, including:
  - conceptual and logical design of a database
  - the query language SQL and the relational algebra
- the design of data-intensive web applications, including:
  - ways to interact with a DBMS
  - the approach “MVC-2 Servlet centric”
  - the treatment of multimedia datasets in a DBMS

The course includes also lab practices about:

- the use of PostgreSQL
- implementation of web applications based on Servlet and JSP
Databases for Bioinformatics
Degree in Bioinformatics
Carlo Combi

The class focuses on:
- the design of a database and its applications, including:
  - conceptual and logical design of a database
  - the query language SQL and the relational algebra
- the design of data-intensive web applications, including:
  - ways to interact with a DBMS
  - the MVC model
- the management of bioinformatics information in databases:
  - XML for bioinformatics

The course includes also lab practices about:
- the use of PostgreSQL/PostBio
- implementation of web applications based on Servlet
This class aims at describing the basic concepts and the main design methodologies of the modern communication networks, both local and geographical. Covered topics include:

- **general concepts**: classification (WAN, LAN, PAN), topologies, reference models (ISO-OSI and TCP/IP)
- **focus on the application level**: DNS, the client/server model, FTP, electronic mail (SMTP, POP3 and IMAP), World Wide Web (HTTP, HTTPS, CGI), Simple Network Management Protocol (SNMP), Secure Shell (SSH)
- **focus on the transport level**: TCP, UDP, secure socket layer (SSL)
- **focus on the network level**: the IP protocol, dynamic host configuration protocol (DHCP), Internet control management protocol (ICMP), routing protocols, IPv6
- **focus on the data-link level**: aims and services, framing
- **focus on the IEEE 802.X standards**: topologies and protocols, the MAC sub-layer, the LLC sub-layer, virtual LAN, wireless LAN
The theoretical part of the course introduces the basic scientific and professional notions of software engineering, addressing in particular the different phases of the software development process:

- planning, design, modeling and specification, implementation, testing and validation, evaluation, maintenance

The programming projects carried out in the lab complement the theory lectures with the design and realization of complex software systems "in the large".
Information systems in medicine
Master Degree in Bioinformatics and Medical Biotechnologies
Carlo Combi

The class focuses on:

- temporal modeling, reasoning, and databases
- temporal reasoning and maintenance in medicine:
  - temporal clinical databases and data warehouse/OLAP systems
  - abstraction of time-oriented clinical data
- time in clinical tasks:
  - time in clinical diagnosis
  - automated support to clinical guidelines and care plans
- the display of time-oriented clinical data and knowledge
Advanced databases
Master Degree in Computer Engineering and Computer Science
Carlo Combi

This course provides the main theoretical concepts of

- the relational data model:
  - relational calculus
  - functional dependencies
  - normal forms and decompositions

- temporal databases
  - data models
  - query languages
  - technological issues

- object and object-relational databases
  - data models
  - query languages

- semistructured databases
  - XML
  - XML Schema and DTD
  - XPath and Xquery
Multimedia and geographical information systems
Master Degree in Computer Engineering and Computer Science
Alberto Belussi

The course provides the concepts and methodologies for the design and implementation of spatial and multimedia databases:

- object-oriented databases: ODMG, ODL, and OQL standards
- conceptual modelling of an object oriented databases in UML and mapping a UML class diagram to ODL
- multimedia databases: representation, management and querying multimedia datasets
- geographical databases: spatial data characteristics and spatial relations
- conceptual modeling of a geographical database in GeoUML
- the mappings of a GeoUML conceptual schema onto geo-relational databases
- query languages for geographical databases: a geo-relational algebra
The course provides an introduction to the fundamentals in large-scale parallel computing systems that deals with big data. The course topics cover:

- **programming frameworks**
  - Distributed filesystems (HFS)
  - NoSQL systems (HBase, Cassandra)
  - data and graph processing (MapReduce, Pregel)
  - SQL-like systems (Pig, Hive)

- **algorithmic design**
  - text processing
  - inverted indexing
  - graph analysis

- **datacenter architectures**
  - topologies
  - communication protocols
  - failure management

The course includes also lab practices
The class focuses on:

- **Biomedical databases:**
  - clinical data repository design
  - electronic medical record
  - health information systems
  - interoperability standards and medical classifications

- **Bioinformatics databases:**
  - management of semi-structured data
  - techniques for information retrieval
  - querying of bioinformatics databases
  - XML and bioinformatics
The class focuses on the design and use of information systems within complex organizations:

- **structure and functions of an information system:**
  - economic, organizational, and management issues
  - business process re-engineering

- **cooperative information systems:**
  - workflow systems

- **management information systems**
  - ERP (Enterprise Resource Planning) systems
  - CRM (Customer Relationship Management) systems

- **decision support systems:**
  - data warehouse/OLAP systems
  - data mining techniques

- **Multi-channel Information Systems**
  - cloud
  - mobile
  - analytics/big data
  - social
“Complexity” is the first module of the “Algorithms” course.

The module introduces computational complexity theory:

- computational models
- time complexity
- space complexity
- reductions and completeness
- randomized computational complexity
- approximation algorithms and approximate complexity classes
The class focuses on fundamentals of the information retrieval systems, in particular to text analysis and web retrieval:

- introduction to information retrieval systems, models, and performance evaluation
- text analysis and indexing
- multimedia information retrieval systems
- user interfaces for search and retrieval results display
- web retrieval, web search engines, web spam, search engine optimization (SEO), motori di ricerca semantici, ricerca personalizzata
- XML for publishing
Outline

1. The research group
2. Courses
   - Bachelor courses
   - Master courses
3. Research
   - Research activities
   - Projects
   - Publications
4. Theses and more
Group research summary

Our research activities comprise:
- temporal databases
- spatial databases
- semistructured databases
- information systems on the web
- temporal information systems
- information systems in medicine
- geographical information systems
- process-aware information systems
- large distributed and networking systems
Research interests include:

- temporal functional dependencies
- (temporal) OLAP and data mining on biomedical/bioinformatics data warehouses
- dealing with clinical guidelines through workflow management systems
Research interests include:

- **Conceptual design of geographical databases**
  - models for the conceptual design of geographical databases
  - languages for the specification of spatial integrity constraints
  - comparison of conceptual designs
  - validation of spatial data wrt conceptual designs
  - methods for the updating of geographical databases maintaining topological properties

- **Approximate spatial queries**
  - similarity functions for spatial relations
  - approximate spatial queries
  - optimization methods of approximate spatial queries
  - spatial queries in recommendation systems

- **Robustness of geometric algorithms**
  - Definition of both rounding techniques and techniques based on the minimum distance concept for guaranteeing the robustness of geometric algorithms in a distributed and heterogeneous context
Research interests include:

- Scheduling in large distributed systems
  - Dynamic priority schemes
  - Size based scheduling for parallel systems
- Performance oriented design of networking systems
  - Data center networks
  - Network cache
XML:
- XML and data warehouses
- XML and bioinformatics
- Fuzzy XML

Temporal aspects of:
- semistructured data:
  - definition of temporal data models and query languages
  - definition of constraints for managing time dimensions
- XML information:
  - evolution and versioning of XML documents and schemata
- data warehouses
Temporal aspects of workflows (with C. Combi):
- a proposal of a Time-aware Conceptual Workflow Model (*Temporal Workflow Model*);
- the computational analysis of the controllability of Temporal Workflow Model schemata.

Frameworks for web application development (with A. Belussi):
- a systematical analysis of the most common frameworks w.r.t the data-centric development issues.
- proposal of simple design patterns for data-centric web application development based on frameworks as Liferay or Portlet technology.
External projects and collaborations (I)

- “Studio, progettazione e realizzazione di Enterprise Data Warehouse”
  - C. Combi and S. Degani in collaboration with IBIC s.r.l.

- DW-SAN: Data warehousing, OLAP, data mining e analisi what-if su dati clinico-sanitari con sistemi open source
  - C. Combi, P. Sala, and V. Mijatovic in collaboration with Novaratio s.r.l.

- ”Strumenti di analisi dei dati della Rete Nazionale di Farmacovigilanza (RNFV)””, ”Tecniche OLAP e Data Warehousing in Farmacologia”
  - C. Combi, S. Migliorini, P. Sala, and A. Sabaini with U. Moretti and R. Lora (Dip. di Sanità Pubblica e Medicina di Comunità)

- “Progetto Burundi”
  - C. Combi and G. Pozzani in collaboration with Fondazione pro-Africa and University of Ngozi (Burundi)
External projects and collaborations (II)

- Spatial (approximate) query processing and optimization
  - A. Belussi and S. Migliorini in collaboration with Prof.ssa B. Catania (University Of Genova)

- Conceptual modeling of spatial databases and robustness verification of geometric algorithms
  - A. Belussi and S. Migliorini in collaboration with Prof. G. Pelagatti (Politecnico di Milano)

- “SITAVR: Sistema Informativo Territoriale Archeologico della città di Verona”
  - A. Belussi and S. Migliorini in collaboration with Prof.ssa P. Basso and Dott.ssa P. Grossi (University of Verona, Dip. Spazio tempo immagine e società)

- Analysis of the impact of the size estimation errors in size-based scheduling
  - D. Carra in collaboration with P. Michiardi (Eurecom, France) and M. Dell’Amico (Eurecom, France)
External projects and collaborations (III)

  - B. Oliboni in collaboration with Elisa Quintarelli (Politecnico di Milano)

- XML and time: versioning and evolution of XML schemata
  - B. Oliboni in collaboration with Fabio Grandi (University of Bologna) and Zouhaier Brahmia (University of Sfax, Tunisia)

- Petri Nets Simulated Annealing (PNSA): a genetic algorithm to rectify Petri nets based process models
  - M. Gambini and S. Migliorini in collaboration with Queensland University of Technology (QUT), Brisbane (Australia)

- Comparative analysis of Scientific and Business Workflow Management Systems
  - M. Gambini and S. Migliorini in collaboration with Queensland University of Technology (QUT), Brisbane (Australia)
Publications statistics

Our research work has led to:

- 209 publications in international conferences
- 77 journal papers
- 25 book chapters
- 5 books
- 2 books/proceedings editing
Participation in conference committees

We participate in several international conference committees:

- Conference on Artificial Intelligence in Medicine (AIME)
- ACM International Conference on Information and Knowledge Management (CIKM)
- International Symposium on Temporal Representation and Reasoning (TIME)
- East-European Conference on Advances in Databases and Information Systems (ADBIS)
- International Workshop on Evolution and Change in Data Management (ECDM)
- International Conference and Workshop on Database and Expert Systems Applications (DEXA)
- Flexible Database and Information Systems Technology (FlexDBIST)
- European Symposium on Algorithm
- International European Conference on Parallel and Distributed Computing (Euro-Par)
- International IEEE/Ifip Networking Conference (NETWORKING)
Participation in journal activities

We participate also in several journal activities:

- Information Systems
- Journal of Web Engineering and Technology
- Artificial Intelligence in Medicine
- Data and Knowledge Engineering
- International Journal of Knowledge and Learning
- Methods of Information in Medicine
- Annals of Mathematics and Artificial Intelligence
- ACM Transactions on Autonomous and Adaptive Systems
- IEEE Transactions on Neural Networks
- IEEE/ACM Transaction on Networking
- IEEE Transaction on Parallel and Distributed Systems
- Computer Networks
We are organizing and we will host...

- **TIME 2014 - International Symposium on Temporal Representation and Reasoning**
  - It brings together researchers working on the management of temporal data and reasoning about temporal aspects of information
  - September 8–10, 2014

- **GandALF 2014 - International Symposium on Games, Automata, Logics and Formal Verification**
  - It brings together researchers from academia and industry working in the fields of Games, Automata, Logics, and Formal Verification
  - September 10–12, 2014

- **ICHI 2014 - IEEE International Conference on Healthcare Informatics**
  - It concerns with the application of CS, information systems, and ICT to address problems in healthcare, public health, and everyday wellness
  - September 15–17, 2014
Outline

1. The research group
2. Courses
   - Bachelor courses
   - Master courses
3. Research
   - Research activities
   - Projects
   - Publications
4. Theses and more
Theses percentage: 2004 – Mar 2014

Percentuale tesi per area di ricerca

- AI e robotica (72, 5%)
- Architetture e sistemi (222, 15%)
- Basi di dati (394, 27%)
- Fisica (2, 0%)
- Linguaggi, logica e metodi formali (409, 28%)
- Matematica (129, 9%)
- Modelli di calcolo (23, 2%)
- VIPS (207, 14%)
Theses distribution: 2004 – Mar 2014

Numero tesi per tipo e area di ricerca

Area di ricerca

- Vecchio ordinamento
- Triennale
- Specialistica

AI e robotica
Architetture e sistemi
Basi di dati
Fisica
Linguaggi, logica e metodi formali
Matematica
Modelli di calcolo
VIPS

Numero tesi
12
31
29
13
129
80
63
224
107
85
304
20
95
30
24
139
44
Thesis proposals (I)  
... and some internships

- (Triennali, Posenato) Tesi applicative sulla sperimentazione e personalizzazione del framework Liferay per la realizzazione di applicazioni web
- (Triennali, Belussi) Strumenti innovativi per la progettazione e lo sviluppo di applicazioni web (Portlet, ...)
- (Triennali, Belussi) Progettazione concettuale di basi di dati geografiche: il modello GeoUML
  - analisi e confronto di specifiche
  - modello dei dati
  - modello geometrico: analisi e confronto con altri approcci
  - vincoli di integrità spaziale: analisi e confronto con altri approcci
  - analisi degli strumenti per la gestione di uno schema concettuale GeoUML
- (Triennali, Belussi) Analisi di strumenti per la visualizzazione e gestione del dato geografico
- (Triennali, Belussi) Argomenti vari correlati alle basi di dati
- (Triennali, Carra) Analisi e valutazione delle prestazioni di sistemi NoSQL
Thesis proposals (II)
... and some internships

- (Combi) Modellazione e gestione di workflow temporali in medicina
  - gestione di linee guida
  - gestione di processi sanitari
- (Combi) Modelli computazionali per process-aware information systems
- (Combi) Petri Nets Simulated Annealing (PNSA): evoluzione dell’algoritmo PNSA definito su reti di Petri per la correzione di modelli di processo basati su linguaggi più sofisticati
- (Combi) Analisi OLAP e data mining su dati (temporali) biomedici, nei seguenti ambiti
  - farmacologia
  - immunologia
  - psichiatria
  - emodialisi
  - genetica
Thesis proposals (III)
... and some internships

- (Belussi) Confronto tra modelli concettuali per basi di dati geografiche
- (Belussi) Tecniche per garantire la robustezza nella valutazione delle relazioni topologiche su insiemi di geometrie vettoriali
- (Belussi) Studio e progettazione di strumenti per la visualizzazione di dati geografici guidati dallo schema concettuale
- (Belussi) Studio di tecniche guidate dallo schema concettuale per la migrazione di dati geografici da uno schema fisico all’altro
- (Belussi) Tecniche per la gestione di interrogazioni spaziali approssimate: top-k, skyline basate sulle relazioni di similarità tra relazioni topologiche
- (Belussi, Combi) Supporto al geo-processing distribuito tramite tecnologie di workflow
Thesis proposals (IV)  
... and some internships

- (Carra) Design e valutazione del software di gestione di un datacenter
- (Carra) Algoritmi paralleli in MapReduce
- (Carra) Gestione della memoria in cache di rete
- (Oliboni) Integrazione di XML e data warehouse
- (Oliboni) XML in bioinformatica
- (Oliboni) Aspetti temporali nella gestione di dati XML
- (Oliboni, Combi) Aspetti temporali nella gestione di dati semistrutturati
- (Oliboni, Combi) XML e data mining
- (Posenato) Analisi della complessità computazionale di problemi di controllabilità nell’ambito di modelli di workflow temporali
- (Posenato) Studio e implementazione di algoritmi di controllabilità nell’ambito di modelli di workflow temporali
Possibilità di svolgere esami e tesi presso la Facoltà di Informatica dell’Università di Murcia (Spagna):

- Vari argomenti di lavoro di tesi nell’ambito di:
  - Workflow temporali
  - Sistemi informativi temporali e sanitari

http://www.um.es/
Semistructured Temporal Spatio-geographical Systems

http://stars.di.univr.it